# Secure and dependable: Area-efficient masked and fault-tolerant architectures

Vojtěch Miškovský, Hana Kubátová, Martin Novotný
Czech Technical University in Prague
Faculty of Information Technology
Email: {miskovoj,kubatova,novotnym}@fit.cvut.cz

***Abstract*—Masking is a powerful instrument for protecting cryptographic devices against side-channel analysis. Multiple masking schemes were introduced providing provable security against attacks of arbitrary order even in the presence of glitches. When a device is a part of some safety-critical system, it needs to meet dependability requirements; therefore, it should be protected against spontaneously occurring faults. Existing commonly used fault-tolerance architectures involve high area overhead as so as the masking schemes do. In this paper, we propose architectures meeting dependability properties of simple modular-redundancy schemes and SCA resistance of masking schemes, but decreasing the area overhead utilizing the randomness involved in the masking schemes.**

**We compare our Masked Duplex architecture with Triple Modular Redundancy. While using one less redundant module, our architecture saves around 20% of the area in comparison with TMR in the case of Threshold Implementation of PRESENT cipher, promising more savings for more complex cryptographic schemes.**

***Index Terms*—security, dependability, side-channel analysis, masking, modular redundancy**

## I. Introduction

Safety-critical electronic systems, like those used in the automotive industry or medical devices, need to fulfill strict dependability properties. To ensure the correct and reliable operation of such systems, they need to be designed fault-tolerant. Since these systems are also usually connected to some network, their activity and communication need to be encrypted. To protect the device against malicious activity, we need to introduce appropriate countermeasures. However, as these systems are usually demanded to be small and low-power, we need to design it lightweight. This could be an issue as both dependable architectures and attack countermeasures can introduce high (area, power, or time) overhead. In this paper, we propose a way to mitigate this issue.

### A. Security issues

One of the most serious threats for cryptographic systems is Side-Channel Analysis (SCA). SCA exploits dependency between some secret information and physical characteristics (so-called side channels) of the device, e.g., power consumption [1], [2] or electromagnetic radiation [3], [4]. An attacker can use this dependency to exploit the secret information (e.g., the cipher key).

One of the possible ways to protect the device against SCA is masking [5], [6]. Masking incorporates random masks to mask the intermediate values of the cryptographic algorithm and to eliminate the side-channel leakage. Multiple Boolean masking schemes effective even in the presence of glitches were introduced, like Threshold Implementations [7], Domain-Oriented Masking [8], or others [9]. These schemes are provably secure against attacks of arbitrary order. Nevertheless, these schemes introduce high area and power overhead.

### B. Dependability issues

A safety-critical device needs to be protected against spontaneous faults to ensure the correct operation of the device and fulfill dependability requirements. Conventional methods to protect the device against such faults are modular redundancy architectures like a duplex, Triple-Modular Redundancy (TMR), or N-modular redundancy (NMR) [10], [11]. The main advantage of these architectures is simplicity: Whole cryptographic modules are replicated and supplemented with checkers or majority voters. This approach also does not spoil the security properties of the design [12]. The main disadvantage of this approach is its overhead; for example, more than 200% of additional hardware resources are required to make the design tolerant to at least one fault.

### C. Our contribution

This work aims to propose an architecture based on existing masking schemes and modular redundancy architectures. In comparison with existing methods, this architecture decreases the overhead, while it keeps both the attack-resistance of masking schemes and the simplicity and dependability properties of the modular redundancy architectures. Specifically, we exploit the fact that a fault in the masked circuit causes different (unmasked) outputs when different random masks are used. This property permits us to build less redundant design utilizing the redundancy introduced by the masking scheme itself.

This paper is structured as follows: In Section II, we present principles of masking schemes and modular redundancy architectures. The principles of the proposed architectures using various amounts of redundant modules are described in detail in Section III. A case study of a proposed architecture using Threshold implementation of PRESENT cipher [13] is demonstrated in Section IV. We discuss the reduction of area overhead, and we also provide leakage assessment of our proposed architecture. Finally, the contributions of our work are concluded in Section V.

## II. Related work

In this section, we explain the principles of techniques our work is based on, i.e., Boolean masking and modular redundancy.

### A. Boolean masking

As was stated above, we base our architectures on the randomness of masking. Specifically, we focus on Boolean shared masking schemes [14], [15]. The fundamental principle of these schemes is that each intermediate value $x$ is split into $n$ shares $x_i$, where

$$x = \bigoplus_{i=1}^{n} x_i \tag{1}$$

.

The input of such a scheme is usually produced in a way that the original input (plain text, cipher key) XORed with $n - 1$ random masks serves as the first share while the masks themselves represent the other shares. The output of the scheme is reproduced by XORing all $n$ output shares. Since the masks are generated independently for each encryption, the input, the output, and all intermediate values of the masked encryption are completely random.

A typical example of such a masking scheme is Threshold Implementation [7], which we use in our case study. In Threshold Implementation, the intermediate values are shared according to Equation 1 and each function applied to the shares must meet three properties: correctness, non-completeness, and uniformity.

### B. Modular redundancy

The most straightforward way to increase the dependability is hardware modular redundancy [11]. These schemes are based on simple replication of functional modules. Common examples are duplex architecture, Triple Modular Redundancy (TMR) [10], or N-Modular Redundancy (NMR) [16]. These architectures detect and/or correct both permanent and transient faults of common fault models (stuck-at, bit-flip).

In our work, we offer area-efficient alternatives to TMR and NMR. Similar work called Time-Extended Duplex was presented in [17]. Time-Extended Duplex is an alternative to TMR based on two modules. This architecture is not limited to masked cryptographic algorithms, the area of this architecture for various circuits lies between 90% and 160% of the area of TMR, and the design is quite complicated, including usage of proprietary gates.

TMR consists of three functional modules and three majority voters. Outputs of all three modules are connected to all three majority voters. This architecture can *correct* a single fault. A diagram of this architecture can be seen in Figure 1.

NMR is a generalization of TMR. It consists of $N$ modules and $N$ majority voters. Outputs of all modules are connected to all majority voters. NMR architecture can correct up to $(N - 1)/2$ faults.

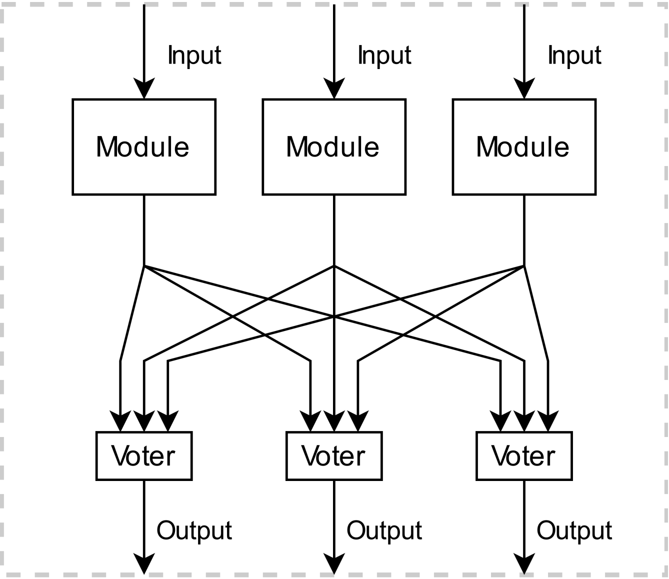


Figure 1: Diagram of a typical TMR architecture

## III. Methodology

To demonstrate our approach, we thoroughly describe an area-efficient architecture — *Masked Duplex* — for correcting one fault, aiming to alternate the TMR architecture. We also propose a generalization of this architecture to achieve an efficient alternative to NMR. Note that in this work, we deal only with fault-tolerance of the data paths. The control unit should be handled independently (as it should be in the case of the traditional modular redundancy architectures).

The architectures rely on the assumption that a fault in the encryption leads to different faulty outputs for the same but differently masked inputs. We assume a round-based implementation of a symmetric cipher as depicted in Figure 2. Such a design can be divided into three parts: input logic (usually just input signals), encryption logic (round logic and round register), and output logic (output signals and, e.g., output multiplexers). Any single fault in the encryption logic leads to different outputs for different random masks with extremely high probability, as such a fault is repeatedly exposed through the whole encryption process. On the other hand, a fault in the input logic or output logic can lead to the same (unshared) outputs for different masks.

### A. Masked Duplex

TMR architecture serves to correct a fault in one of three modules. We propose an architecture with similar properties using two modules only. A diagram of the proposed architecture can be seen in Figure 3. The two modules run the encryption in parallel with the same masks. The outputs of encryption are compared, and if they differ (one of the modules is faulty), the encryption is repeated with new masks (again, the same for both modules). Unmasked outputs of both consecutive encryptions are compared for each of the modules. If consecutive outputs of one of the modules differ,

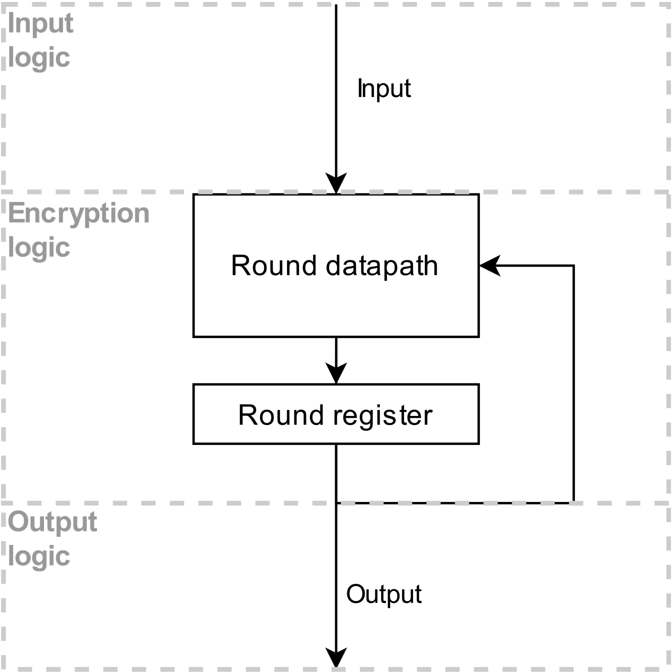


Figure 2: Diagram of round-based cipher implementation

that module is considered faulty, and the output of the other one is propagated. If a fault occurs in the input or output logic, there is a possibility that consecutive outputs do not differ for either of the modules. In such a case, the encryption must be repeated again using new random masks.

As mentioned above, the proposed architecture enables the correction of a single fault, similarly to TMR. Compared to TMR, we spared one of three modules. The encryption takes the same time like in TMR unless a fault is introduced in one of the modules. Also, faults in both modules at the same time can be detected with no additional logic.

*1) Comparison logic — original approach (unmasked outputs):* To protect the design against a fault in the comparison logic, this needs to be triplicated similarly as in the case of TMR. A diagram of an example of comparison logic is in Figure 4. This logic is more complex; therefore, more area demanding than majority voters in TMR. An advantage of the proposed comparison logic lies in area efficiency, as the outputs are unmasked; therefore, a shorter comparator is used. The comparator is multiplexed. It is used for the comparison of outputs of both modules as well as for comparison of outputs of two consecutive encryptions.

*2) Comparison logic — alternative approach (masked outputs):* Unmasked outputs could entail security issues (e.g., for combined side-channel and fault attacks [18], [19]). This problem can be solved by comparing the masked outputs of the modules. A diagram of an alternative comparison logic is in Figure 5. When a fault is detected (the modules produce different outputs), random data are used as the input of the encryption. Repeating the encryption in the same way as in the case of the original comparison logic (but using random data), the faulty module is identified, and the other one can be used to encrypt the original input data. This method demands wider comparators (fitting the masked outputs) and additional time overhead when a fault occurs in comparison with the original comparison logic.

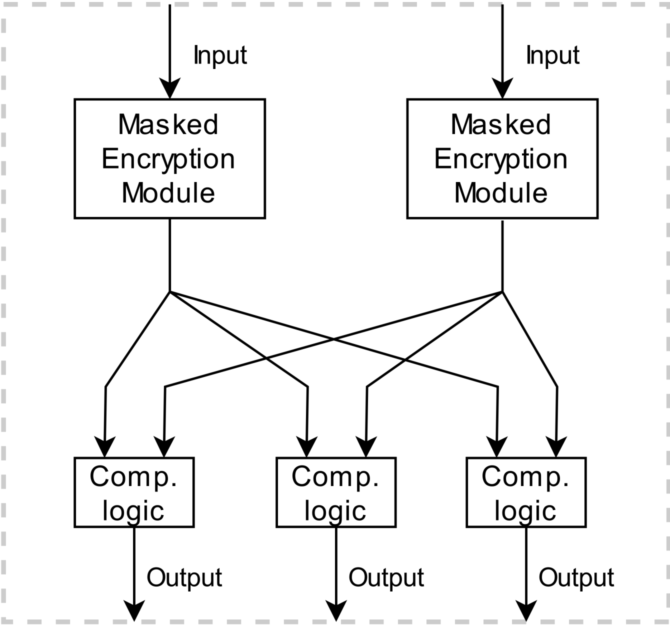


Figure 3: Diagram of proposed Masked Duplex

### B. Generalization of Masked Duplex

As NMR is a generalization of TMR, we can also generalize our Masked Duplex. NMR is tolerant to faults in $n = (N-1)/2$ modules; therefore, $N = 2 \times n + 1$ modules are needed to tolerate $n$ faults. Using the principle proposed in the previous section, we only need one faultless module at a time. If all outputs are not the same, the correct module can be detected by repeating the encryption in the same manner as in the previous subsection. Therefore, we only need $n + 1$ modules to tolerate $n$ faults. Nevertheless, we still need $2 \times n + 1$ comparator circuits that are more complex than the majority voters.

## IV. Case study

In this section, we evaluate the Masked Duplex proposed in Section III-A using three-share Threshold Implementation of the PRESENT cipher [13] as is described in [20]. The cipher is implemented in Spartan-6 FPGA on Sakura-G evaluation board [21]. Four architectures employing TI of PRESENT are compared regarding the area of the design: single module, TMR, and our Masked Duplex using either the original or the alternative comparison logic. The power leakage of all three implementations is also evaluated.

### A. Results

All three designs were synthesized using Xilinx XST (Xilinx ISE 14.7). A comparison of slice utilization for each of the implementations can be seen in Table I. Our architecture

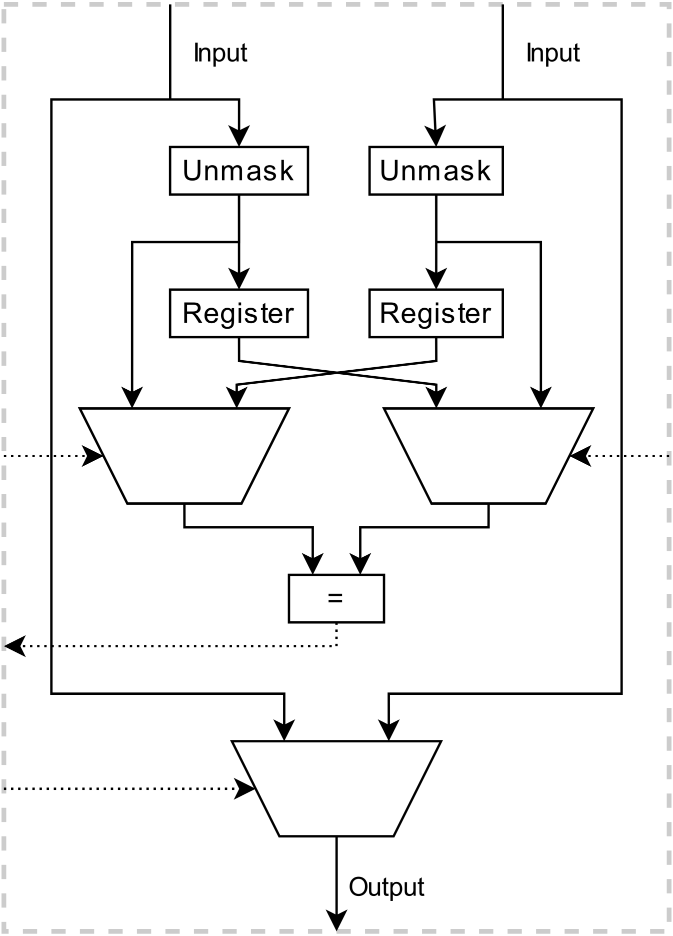


Figure 4: Diagram of the proposed comparison logic for Masked Duplex

Table I: Comparison of area demands of evaluated implementations

| Design | Slice utilization | Overhead |
|---|---|---|
| Single module | 2199 | 0% |
| TMR | 7180 | 227% |
| Masked Duplex (orig.) | 5764 | 162% |
| Masked Duplex (alt.) | 6589 | 200% |

is around 20% smaller than standard TMR (in case of the original comparison logic). As the PRESENT is a light-weight cipher, even higher savings can be expected for more complex ciphers like AES [22]. For such a more area-demanding encryption module, the ratio between the area saved by the spared encryption module and the area increased by the more complex comparison logic would be higher and, therefore, the advantage of our approach would be even more significant.

### B. Leakage Assessment

Leakage Assessment was performed for each implementation. The leakage was evaluated for 1,000,000 traces using non-specific, fixed vs. random, first-order Welch's t-test [23]. The t-test enables us to verify whether samples in two sets

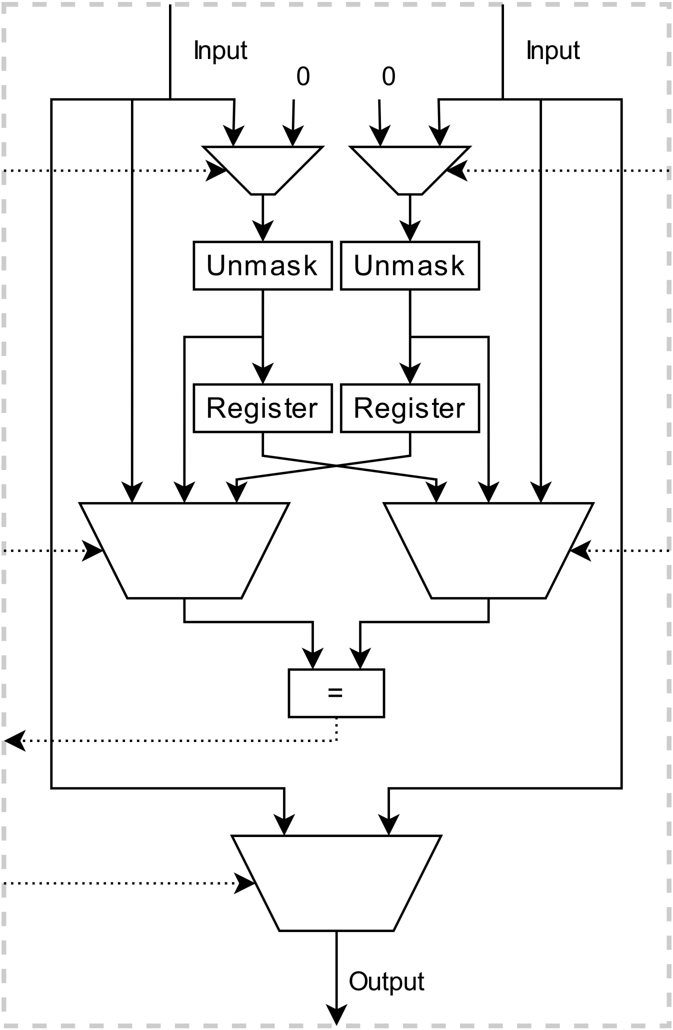


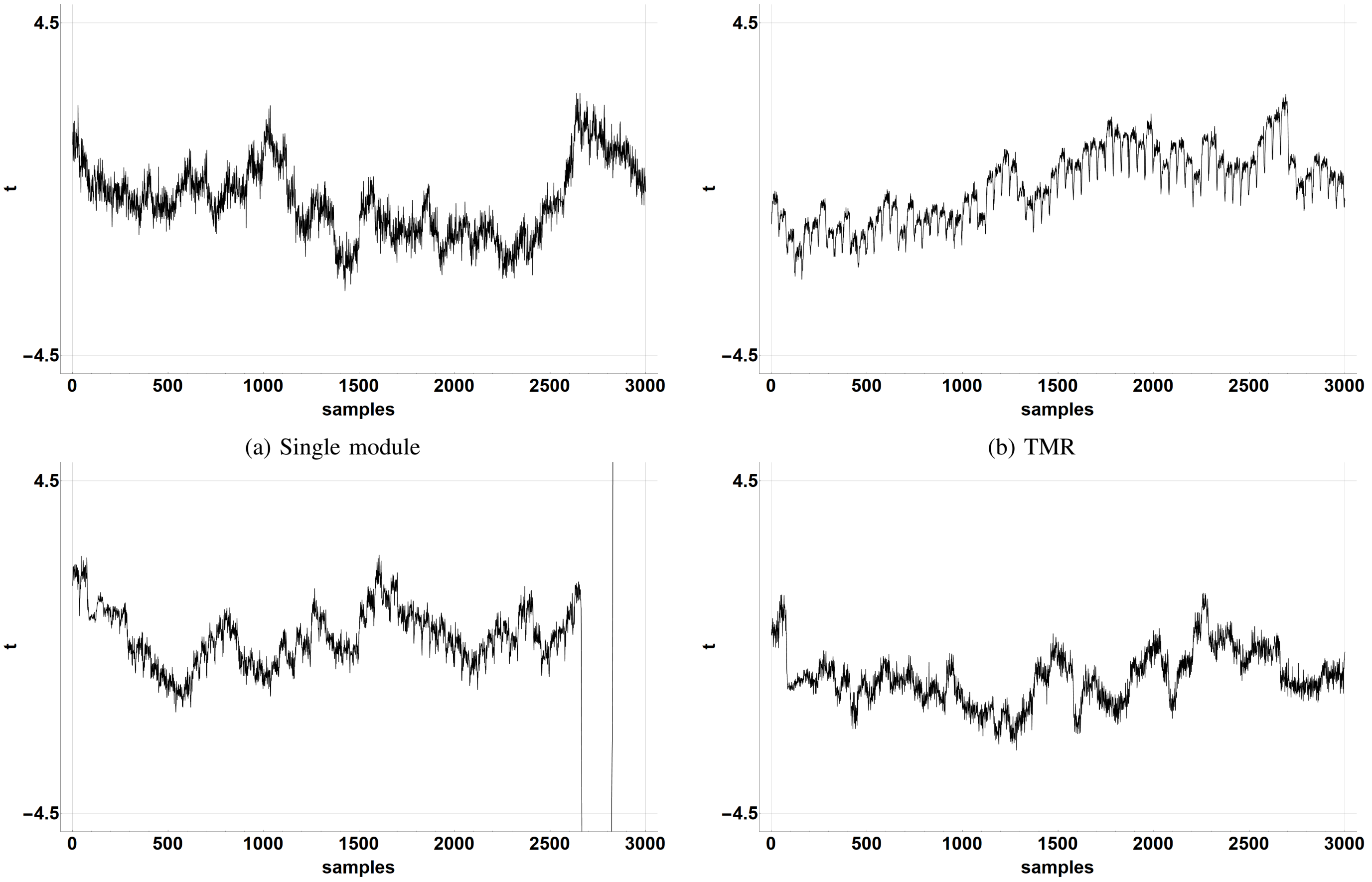
Figure 5: Diagram of the alternative comparison logic for Masked Duplex

were drawn from the same population. In our case, we compare sets of power traces obtained during the encryption of fixed data and random data. For each sample point, the t-test statistic $t$ is computed as

$$t = \frac{\mu_0 - \mu_1}{\sqrt{\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}}}$$

, where $\mu_0$ and $\mu_1$ are sample means, $s_0^2$ and $s_1^2$ are sample variances, and $n_0$ and $n_1$ are cardinalities of each set. The value 4.5 is usually considered a threshold for the t-value to reject the hypothesis that the two sets were drawn from the same population. Therefore, when the t-value for all the sample points lies within the interval $(-4.5, 4.5)$, no leakage is detected.

Plots of t-values can be seen in Figure 6. As we can see, there is no leakage in the case of a single module (6a) and TMR (6b). In the case of Masked Duplex with the original comparison logic (6c), there is significant leakage at the end of the encryption. This leakage is caused by the unmasked



(a) Single module

(b) TMR

(c) Masked Duplex (orig.)

(d) Masked Duplex (alt.)

Figure 6: Plots of t-values

outputs in the comparison logic, as it is discussed in Section III-A. This leakage is successfully eliminated by the alternative comparison logic (6d).

## V. Conclusion

In this paper, we dealt with the issue of having a device both secure and dependable at the same time. We proposed novel architectures based on masking schemes exploiting the involved randomness. These architectures keep the simplicity of modular redundancy, but they decrease the number of redundant modules; therefore, they significantly decrease the area overhead. In the case of TI of PRESENT cipher, the area overhead of standard TMR architecture is 227%, while the area overhead of the Masked Duplex is only 162%. The savings would be even more significant for some more complex cryptographic algorithm. Implemented architecture also passed the Welch's t-test with an exception of leakage of the comparison logic, where the outputs appear unmasked. We proposed an alternative comparison logic eliminating this leakage. This alternative comparison logic provides a possible trade-off between the security and the overhead of the proposed architectures.

Considering these facts, we can conclude that we proposed a novel approach for secure and dependable design, exploiting the redundancy introduced by a masking scheme, with similar qualities and lower overhead in comparison with existing methods. The results of the proof of concept presented in this paper encourage us to a deeper investigation of presented architectures, including various cryptographic algorithms, various masking schemes, and higher numbers of modules.

## VI. Acknowledgment

This research has been partially conducted during author's Erasmus internship at Ruhr-Universitaet Bochum under supervision of prof. Amir Moradi. It has been partially supported by the grant GA16-05179S of the Czech Grant Agency "Fault-Tolerant and Attack-Resistant Architectures Based on Programmable Devices: Research of Interplay and Common Features" (2016-2018); and CTU project SGS17/213/OHK3/3T/18.

## References

[1] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Annual International Cryptology Conference*, pp. 388–397, Springer, 1999.

[2] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 16–29, Springer, 2004.

[3] J.-J. Quisquater and D. Samyde, "Electromagnetic analysis (ema): Measures and counter-measures for smart cards," *Smart Card Programming and Security*, pp. 200–210, 2001.

[4] J. R. Rao and P. Rohatgi, “EMpowering Side-Channel Attacks.,” *IACR Cryptology ePrint Archive*, vol. 2001, p. 37, 2001.

[5] S. Chari, C. S. Jutla, J. R. Rao, and P. Rohatgi, “Towards sound approaches to counteract power-analysis attacks,” in *Annual International Cryptology Conference*, pp. 398–412, Springer, 1999.

[6] J. Blömer, J. Guajardo, and V. Krummel, “Provably secure masking of AES,” in *International Workshop on Selected Areas in Cryptography*, pp. 69–83, Springer, 2004.

[7] S. Nikova, C. Rechberger, and V. Rijmen, “Threshold implementations against side-channel attacks and glitches,” in *International Conference on Information and Communications Security*, pp. 529–545, Springer, 2006.

[8] H. Groß, S. Mangard, and T. Korak, “Domain-oriented masking: Compact masked hardware implementations with arbitrary protection order.,” in *TIS@ CCS*, p. 3, 2016.

[9] O. Reparaz, B. Bilgin, S. Nikova, B. Gierlichs, and I. Verbauwhede, “Consolidating masking schemes,” in *Annual Cryptology Conference*, pp. 764–783, Springer, 2015.

[10] R. E. Lyons and W. Vanderkulk, “The use of triple-modular redundancy to improve computer reliability,” *IBM journal of research and development*, vol. 6, no. 2, pp. 200–209, 1962.

[11] I. Koren, *Fault-tolerant systems*. Amsterdam Boston: Elsevier/Morgan Kaufmann, 2007.

[12] J. Říha, V. Miškovský, H. Kubátová, and M. Novotný, “Influence of fault-tolerance techniques on power-analysis resistance of cryptographic design,” in *2017 Euromicro Conference on Digital System Design (DSD)*, pp. 260–267, IEEE, 2017.

[13] A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. Robshaw, Y. Seurin, and C. Vikkelsoe, “Present: An ultra-lightweight block cipher,” in *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 450–466, Springer, 2007.

[14] J.-S. Coron and L. Goubin, “On boolean and arithmetic masking against differential power analysis,” in *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 231–237, Springer, 2000.

[15] J.-S. Coron, J. Großschädl, and P. K. Vadnala, “Secure conversion between boolean and arithmetic masking of any order,” in *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 188–205, Springer, 2014.

[16] I. Koren and S. Y. H. Su, “Reliability analysis of n-modular redundancy systems with intermittent and permanent faults,” *IEEE Transactions on Computers*, no. 7, pp. 514–520, 1979.

[17] J. Belohoubek, P. Fiser, and J. Schmidt, “Novel c-element based error detection and correction method combining time and area redundancy,” in *2015 Euromicro Conference on Digital System Design*, pp. 280–283, IEEE, 2015.

[18] F. Amiel, K. Villegas, B. Feix, and L. Marcel, “Passive and active combined attacks: Combining fault attacks and side channel analysis,” in *Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC 2007)*, pp. 92–102, IEEE, 2007.

[19] T. Roche, V. Lomné, and K. Khalfallah, “Combined fault and side-channel attack on protected implementations of aes,” in *International Conference on Smart Card Research and Advanced Applications*, pp. 65–83, Springer, 2011.

[20] A. Poschmann, A. Moradi, K. Khoo, C.-W. Lim, H. Wang, and S. Ling, “Side-channel resistant crypto for less than 2,300 ge,” *Journal of Cryptology*, vol. 24, no. 2, pp. 322–345, 2011.

[21] H. Guntur, J. Ishii, and A. Satoh, “Side-channel attack user reference architecture board sakura-g,” in *Consumer Electronics (GCCE), 2014 IEEE 3rd Global Conference on*, pp. 271–274, IEEE, 2014.

[22] N. F. Pub, “197: Advanced encryption standard (AES),” *Federal Information Processing Standards Publication*, vol. 197, no. 441, p. 0311, 2001.

[23] T. Schneider and A. Moradi, “Leakage assessment methodology,” in *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 495–513, Springer, 2015.